
ECONOMETRIA II

(2001/2002)

Modelização Univariada de Séries Temporais: uma Introdução

Artur C. B. da Silva Lopes

Instituto Superior de Economia e Gestão

Universidade Técnica de Lisboa

Nota Introdutória

Estas notas foram elaboradas com o propósito de apoiar o estudo dos alunos da disciplina de Econometria II do ISEG, no ano lectivo de 2001/2002, relativamente à matéria do capítulo 4 do programa. A principal fonte empregue foi o livro de Jonhston, J. e DiNardo, J. (1997), *Econometric Methods*, 4th ed. [JD], mas procurou-se simplificar a abordagem aí seguida. No entanto, por vezes este texto segue esse livro de perto. O leitor interessado também poderá consultar, entre muitas outras, as seguintes obras adicionais: Franses, Philip Hans (1998), *Time Series Models for Business and Economic Forecasting*, Mills, Terence C. (1992), *Time Series Techniques for Economics* e Mills, Terence C. (1999), *The Econometric Modelling of Financial Time Series*, 2nd ed. . Todos estes livros são da Cambridge University Press. Também o texto de Cochrane, J. H. (1997), *Time Series for Macroeconomics and Finance*, disponível (em formato PDF) em <http://gsbwww.uchicago.edu/fac/john.cochrane.research/Papers/timeser1.pdf>, embora não contendo uma abordagem estatística, é recomendado pela sua acessibilidade.

Agradeço os comentários da colega Isabel Proença. Comentários adicionais, sugestões e críticas serão muito bem-vindos!

Lisboa, 25 de Abril de 2002
Artur C. B. da Silva Lopes

Índice

1	Introdução	4
2	Preliminares	5
2.1	O operador de desfasamento (<i>lag operator</i>)	5
2.2	O operador de diferenciação	6
2.3	Polinómios em L	6
2.4	Modelização ARMA	6
3	Séries Estacionárias e Não Estacionárias	7
3.1	O Caso do AR(1)	8
3.2	Raiz Unitária	9
3.3	Análise Gráfica	10
4	Modelos ARMA para Séries Estacionárias	13
5	Previsão	16
6	Modelos Autoregressivos	19
6.1	Estimação	20
6.2	Seleção	20
6.3	Testes-diagnóstico	21
6.4	Exemplo	22
7	Análise de Estacionaridade	24
7.1	Séries económicas: estacionárias em tendência (TSP) ou por diferenci- ação (DSP)?	25
7.2	Testes de raízes unitárias: Testes DF e ADF	29
7.2.1	Testes DF	29
7.2.2	Testes ADF	32
7.3	Exemplo	33

1 Introdução

Neste capítulo o estudo incide exclusivamente sobre a modelização de séries temporais económicas, isto é, de variáveis económicas observadas ao longo do tempo. Mais ainda: trata-se apenas da chamada modelização **univariada** ou **extrapolativa**.

Diz-se univariada porque se despreza a possibilidade de existirem relações entre a variável de interesse (y_t) e outras variáveis económicas. Assim, o comportamento de y_t é explicado apenas com base na sua história passada e nos valores corrente e passados de um erro estocástico, assumido evoluir segundo um processo ruído branco ($\{\epsilon_t\}$):

$$y_t = f(y_{t-1}, y_{t-2}, \dots, \epsilon_t, \epsilon_{t-1}, \dots), \quad \epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2).$$

Para tornar operacional esta relação é necessário especificar: a) a sua forma funcional e b) os números de desfasamentos (*lags*) de y_t e de ϵ_t . Relativamente ao primeiro aspecto saliente-se que só consideraremos modelos lineares (mas note-se que o estudo dos modelos não-lineares está em franca expansão). Relativamente ao segundo, de acordo com a sua estrutura e o seus desfasamentos teremos vários modelos, já apresentados no capítulo 2:

$$\text{AR}(1) : \quad y_t = \phi y_{t-1} + \epsilon_t,$$

$$\text{AR}(p) : \quad y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t,$$

$$\text{MA}(1) : \quad y_t = \epsilon_t - \theta \epsilon_{t-1},$$

$$\text{MA}(q) : \quad y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}.$$

Como é sabido, todos estes modelos podem ser vistos como casos particulares do modelo (misto) ARMA(p, q):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}.$$

Note-se ainda que todos estes modelos podem ser aumentados através da introdução de variáveis determinísticas: constante, tendência, *dummies* sazonais, etc..

Porquê modelização extrapolativa? Porque o principal objectivo destes modelos é o de prever os valores futuros das variáveis extrapolando o seu comportamento passado. Contudo, note-se que esta extrapolação é algo sofisticada, pois baseia-se em métodos estatísticos.

Porquê e quando utilizar este tipo de abordagem? Como se disse, este tipo de modelização é puramente estatístico e despreza a informação dada pela teoria económica sobre a relação da variável com outras. Isto faz algum sentido? Não é a modelização multivariada, que toma em consideração essas relações, preferível?

Não necessariamente. Porquê?

- a) Porque quando o objectivo é a previsão de **curto prazo**, estes modelos são bastante “baratos” (porque dispensam a obtenção de informação sobre outras variáveis) e a sua relação qualidade/preço pode ser (quase) imbatível. Na verdade, modelos simples deste tipo podem fornecer melhores previsões de curto prazo do que modelos com muitas equações e muitas variáveis explicativas.
- b) Porque a relação com outras variáveis pode não estar bem fundamentada teoricamente.

Adicionalmente, como veremos mais adiante, a boa modelização multivariada de séries temporais requer que tenha sido efectuada **previamente** uma análise univariada de cada série.

2 Preliminares

2.1 O operador de desfasamento (*lag operator*)

Aplicado a uma série, o operador de desfasamento, L , atrasa-a ¹ ou desfasa-a num período:

$$L(y_t) = y_{t-1}.$$

Assim, $L^2(y_t) = L[L(y_t)] = L(y_{t-1}) = y_{t-2}$. Mais geralmente, com s inteiro:

$$L^s(y_t) = y_{t-s}.$$

Note-se que este operador é:

- comutativo com a multiplicação: $L(\beta y_t) = \beta L(y_t) = \beta y_{t-1}$;
- distributivo relativamente à adição: $L(y_t + x_t) = y_{t-1} + x_{t-1}$.

Note-se ainda que, embora nas manipulações algébricas o operador L possa ser tratado como um escalar, não se deve perder de vista que é um operador.

¹Na literatura tradicional de séries temporais é mais comum a utilização de B , como operador de atraso (*backward shift*).

2.2 O operador de diferenciação

O operador de diferenciação, Δ , pode escrever-se como $1 - L$, isto é, $\Delta \equiv 1 - L$. De facto,

$$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t.$$

Repare-se que também pode aparecer sob a forma de potência: $\Delta^2 y_t = \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = y_t - y_{t-1} - y_{t-1} + y_{t-2} = y_t - 2y_{t-1} + y_{t-2}$.

2.3 Polinómios em L

Um polinómio simples em L é, por exemplo, $(1 - L)L$, isto é, ΔL . Assim,

$$(1 - L)Ly_t = \Delta y_{t-1} = y_{t-1} - y_{t-2}.$$

Um exemplo mais interessante é o seguinte:

$$y_t + 0.8y_{t-1} - 0.2y_{t-2} + 0.3y_{t-3} = (1 + 0.8L - 0.2L^2 + 0.3L^3)y_t.$$

Fazendo $\phi(L) = (1 + 0.8L - 0.2L^2 + 0.3L^3)$, podemos escrever

$$y_t + 0.8y_{t-1} - 0.2y_{t-2} + 0.3y_{t-3} = \phi(L)y_t.$$

Uma operação importante com os polinómios em L é a da inversão. O caso que se apresenta a seguir, embora simples, é dos mais importantes. Seja agora $\phi(L) = 1 - \phi L$. Repare-se então que

$$(1 - \phi L)(1 + \phi L + \phi^2 L^2 + \dots + \phi^p L^p) = 1 - \phi^{p+1} L^{p+1}.$$

E desde que $|\phi| < 1$, quando $p \rightarrow \infty \Rightarrow \phi^{p+1} L^{p+1} \rightarrow 0$. Então, o polinómio inverso de $\phi(L)$ é

$$\phi^{-1}(L) = \frac{1}{1 - \phi L} = 1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \dots,$$

isto é, um polinómio infinito. Uma forma intuitiva de ver isto consiste em olhar para o terceiro termo da igualdade como uma série geométrica com primeiro termo 1 e razão ϕL .

2.4 Modelização ARMA

A modelização (e previsão) ARMA consiste basicamente em 3 passos:

1. Análise de estacionaridade da série. Muito frequentemente é necessário transformar a série (logaritmizá-la e/ou diferenciá-la) para a “estacionarizar”. Saliente-se que os modelos ARMA pressupõem a estacionaridade da série ².
2. Com base na análise das propriedades de autocorrelação da série (eventualmente transformada) escolhemos algum ou alguns modelos que a possam representar aproximadamente. Trata-se da chamada “**análise de identificação**”. No caso de considerarmos vários modelos alternativos procede-se à selecção de um modelo com base em “testes-diagnóstico” e em estatísticas como a AIC ou a SC. Note-se que o estudo que faremos desta fase será muito limitado pois apenas consideraremos modelos autoregressivos. Por outro lado, a escolha de um modelo bom e simples é, em boa medida, uma arte, requerendo muita experiência.
3. O modelo seleccionado é utilizado para efectuar previsão.

3 Séries Estacionárias e Não Estacionárias

Já sabemos que para que uma série temporal (ou melhor, um processo estocástico) seja estacionária(o) (fracamente ou em covariância), devem ser satisfeitas as seguintes condições:

1. $E(y_t) = \mu, t = 0, 1, 2, \dots,$
2. $E[(y_t - \mu)(y_{t-j} - \mu)] = \gamma_j, t = 0, 1, \dots, j = 0, 1, 2, \dots$

Mas note-se que fazendo $j = 0$ nesta igualdade resulta

$$E[(y_t - \mu)^2] = \sigma_y^2 = \gamma_0, t = 0, 1, 2, \dots$$

Assim, uma série temporal é estacionária se a sua média, a sua variância e as suas autocovariâncias são independentes do tempo, isto é, são constantes ao longo do tempo. Note-se que esta definição de estacionaridade se baseia apenas nos dois primeiros momentos do processo estocástico.

²Os modelos para séries que não são estacionárias mas que se tornam estacionárias quando diferenciadas serão chamados de ARIMA. O “I” abrevia a palavra “integrated”.

3.1 O Caso do AR(1)

A título de exemplo considere-se o processo AR(1):

$$y_t = m + \phi y_{t-1} + \epsilon_t.$$

Substituindo y_{t-1} pela sua expressão, isto é, por $m + \phi y_{t-2} + \epsilon_{t-1}$ tem-se

$$y_t = m + \phi m + \phi^2 y_{t-2} + \epsilon_t + \phi \epsilon_{t-1}.$$

Continuando a resolver recursivamente e supondo que o processo se iniciou há muito tempo atrás tem-se:

$$y_t = m(1 + \phi + \phi^2 + \phi^3 + \dots) + (\epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots), \quad (1)$$

isto é, o processo AR(1) escrito sob a forma de MA(∞). Aplicando o operador valor esperado a esta expressão:

$$E(y_t) = m(1 + \phi + \phi^2 + \phi^3 + \dots).$$

Desta forma, o processo tem média (finita) sse a série (geométrica) dentro de parêntesis for convergente. Como é sabido, a condição necessária e suficiente é que $|\phi| < 1$. Nestas condições,

$$E(y_t) = \mu = \frac{m}{1 - \phi}, \quad (2)$$

independente de t .

Para calcular a variância do processo vai continuar a assumir-se que $|\phi| < 1$. Assim, note-se que $y_t - \mu = \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots$. Por conseguinte,

$$\sigma_y^2 = Var(y_t) = \gamma_0 = E[(y_t - \mu)^2] = E[\epsilon_t^2 + \phi^2 \epsilon_{t-1}^2 + \phi^4 \epsilon_{t-2}^2 + \dots + 2\phi \epsilon_t \epsilon_{t-1} + \dots],$$

que também existe finita sse $|\phi| < 1$. Nestas condições,

$$\sigma_y^2 = Var(y_t) = \gamma_0 = \sigma_\epsilon^2 + \phi^2 \sigma_\epsilon^2 + \phi^4 \sigma_\epsilon^2 + \dots = \frac{\sigma_\epsilon^2}{1 - \phi^2},$$

e note-se que esta expressão não faria qualquer sentido se $|\phi| \geq 1$. Note-se também que, desde que a condição $|\phi| < 1$ seja satisfeita, a variância do processo não depende de t .

Finalmente, pode ainda mostrar-se que, desde que $|\phi| < 1$, as autocovariâncias são dadas por: $\gamma_1 = \phi \sigma_y^2$, $\gamma_2 = \phi^2 \sigma_y^2$ e, em geral, $\gamma_s = \phi^s \sigma_y^2$, $s = 0, 1, 2, \dots$, que são

independentes de t (dependendo apenas da ordem de desfasamento, s). Em resumo, quando $|\phi| < 1$ o processo AR(1) é estacionário ³.

Refira-se ainda que como os coeficientes de autocorrelação são dados por $\rho^s = \gamma_s/\gamma_0$, tem-se $\rho^s = \phi^s, s = 0, 1, 2, \dots$. Desta forma, um processo AR(1) estacionário tem “memória longa”, e tanto mais longa quanto mais próximo de $|1|$ estiver ϕ . Por exemplo, com $\phi = 0.9$, $\rho^{20} \approx 0.122$. Isto é, mesmo após 20 períodos, a série ainda se lembra de 12.2% do choque ocorrido.

3.2 Raiz Unitária

Considere-se de novo o processo AR(1) mas com $\phi = 1$, caso em que se tem o **passeio aleatório com deriva** ($m \neq 0$):

$$y_t = m + y_{t-1} + \epsilon_t.$$

Diz-se que este processo tem uma raiz unitária pois ele pode escrever-se como $(1 - L)y_t = m + \epsilon_t$ e a equação característica, $\phi(z) = 0$, ⁴ é dada por $1 - z = 0$, com solução $z = 1$. Ou seja, tem-se uma raiz de módulo unitário (e que é mesmo igual a um).

Da análise anterior já sabemos que este processo não é estacionário. Porquê? Porque nem sequer tem média constante. De facto, resolvendo recursivamente e assumindo que o processo se iniciou no momento 0 tem-se

$$y_t = 2m + y_{t-2} + \epsilon_t + \epsilon_{t-1},$$

...

$$y_t = y_0 + mt + \sum_{i=1}^t \epsilon_i. \quad (3)$$

Saliente-se a presença de um termo de tendência determinística (mt). Refira-se ainda que ao termo de acumulação de todos os choques, desde o primeiro período até ao presente ($\sum_{i=1}^t \epsilon_i$), é usual chamar de **tendência estocástica**. Desta forma, o valor esperado condicional,

$$E(y_t|y_0) = y_0 + mt,$$

varia ao longo do tempo.

³Recorde-se que $|\phi| < 1$ é também a condição de estabilidade da equação às diferenças de primeira ordem $y_t = m + \phi y_{t-1}$. Desta forma também se pode dizer que, desde que a condição seja satisfeita, o processo AR(1) é estável.

⁴É usual empregar z em vez de L na equação característica pois L é um operador.

Por outro lado, para o **passeio aleatório sem deriva** ($m = 0$),

$$x_t = x_{t-1} + \epsilon_t$$

o problema reside sobretudo na variância. De facto, resolvendo recursivamente tem-se $x_t = x_{t-2} + \epsilon_t + \epsilon_{t-1}, \dots, x_t = x_0 + \sum_{i=1}^t \epsilon_i$ (note-se, de novo, a presença da tendência estocástica). Por conseguinte,

$$\text{Var}(x_t|x_0) = \text{Var}\left(\sum_{i=1}^t \epsilon_i\right) = t\sigma_\epsilon^2,$$

que é claramente não constante, crescendo à medida que o tempo passa.

Ambos os passeios aleatórios são ditos **processos integrados de primeira ordem**, escrevendo-se $y_t \sim I(1)$ e $x_t \sim I(1)$. De facto, trata-se dos exemplos mais simples de processos integrados de primeira ordem. Porquê integrados? Porquê de primeira ordem? Porque são não estacionários mas quando são diferenciados uma vez ficam estacionários:

$$\Delta y_t = y_t - y_{t-1} = (m + \epsilon_t) \sim I(0),$$

$$\Delta x_t = x_t - x_{t-1} = \epsilon_t \sim I(0),$$

onde $I(0)$ é a notação usada para representar uma série que não precisa ser diferenciada nenhuma vez para ficar estacionária (mas que pode ser não estacionária).

Assim, **ordem de integração** de um processo é o número mínimo de vezes que é necessário diferenciá-lo para o tornar estacionário.

3.3 Análise Gráfica

No primeiro gráfico da “figure 1” encontra-se representada graficamente uma série, gerada em computador com base num modelo AR(1) estacionário, dado por $x_t = 0.7x_{t-1} + \epsilon_t$ (com $x_0 = 0$). Ou seja, neste caso a média é nula.

Atente-se nas suas principais características:

- o comportamento da série é pouco suave ou muito errático;
- a série tem um comportamento de **reversão** (ou regressão) **para a média**, cruzando frequentemente o eixo das abcissas ⁵;

⁵Se se tivesse $m \neq 0$ a média da série seria dada por $m/(1 - 0.7)$ e a série cruzaria frequentemente a recta $y = m/(1 - 0.7)$. Note-se que um ϕ mais próximo de zero implicaria um comportamento de reversão para a média ainda mais acentuado.

- como a série tem variância (finita e) constante, ela tende a flutuar sempre dentro de um intervalo de amplitude constante.

Por outro lado, no segundo gráfico da “figure 1” apresenta-se um exemplo de um passeio aleatório sem deriva, $y_t = y_{t-1} + \epsilon_t$ (com $y_0 = 0$). Repare-se nas seguintes características:

- tem um comportamento muito mais suave;
- não apresenta um comportamento de reversão para a média nem para nenhum valor constante (isto é, a série parece vaguear sem destino);
- como a variância cresce com o tempo, a série tende a flutuar dentro de intervalos de amplitude crescente.

Finalmente, no terceiro gráfico apresenta-se um exemplo de um processo AR(1) não estacionário explosivo (não integrado), dado por $w_t = 1.05w_{t-1} + \epsilon_t$ (isto é, $|\phi| > 1$). Séries económicas com gráficos deste tipo são usualmente logaritmizadas na tentativa de estacionarização. A série assim transformada é então sujeita a uma análise de estacionaridade.

Porque não considerar também séries (estacionárias e não estacionárias) com coeficiente autoregressivo negativo? Porque estas têm pouco interesse em economia. De facto, as séries económicas são geralmente caracterizadas por um comportamento de inércia: a valores elevados tendem a seguir-se valores elevados e a valores baixos tendem a suceder-se também valores baixos. Por exemplo, é muito difícil (se não impossível) encontrar uma série económica bem aproximada pelo modelo não estacionário $\eta_t = -\eta_{t-1} + \epsilon_t$, que também possui uma raiz de módulo unitário ($z = -1$). De facto, repare-se que a série tenderia a alternar de sinal, de observação para observação. Algumas séries económicas com sazonalidade podem apresentar este tipo de comportamento mas só depois de diferenciadas. Ou seja, o interesse recai sobretudo sobre a raiz unitária positiva.

O principal problema a estudar aqui será então o de, por exemplo, no contexto do modelo AR(1), $y_t = m + \phi y_{t-1} + \epsilon_t$, testar

$$H_0 : \phi = 1 \text{ vs } H_1 : \phi < 1.$$

Aparentemente, o problema parece ter solução simples: estima-se a equação com o OLS e emprega-se a estatística— t

$$\frac{\hat{\phi} - 1}{\hat{\sigma}_{\hat{\phi}}}.$$

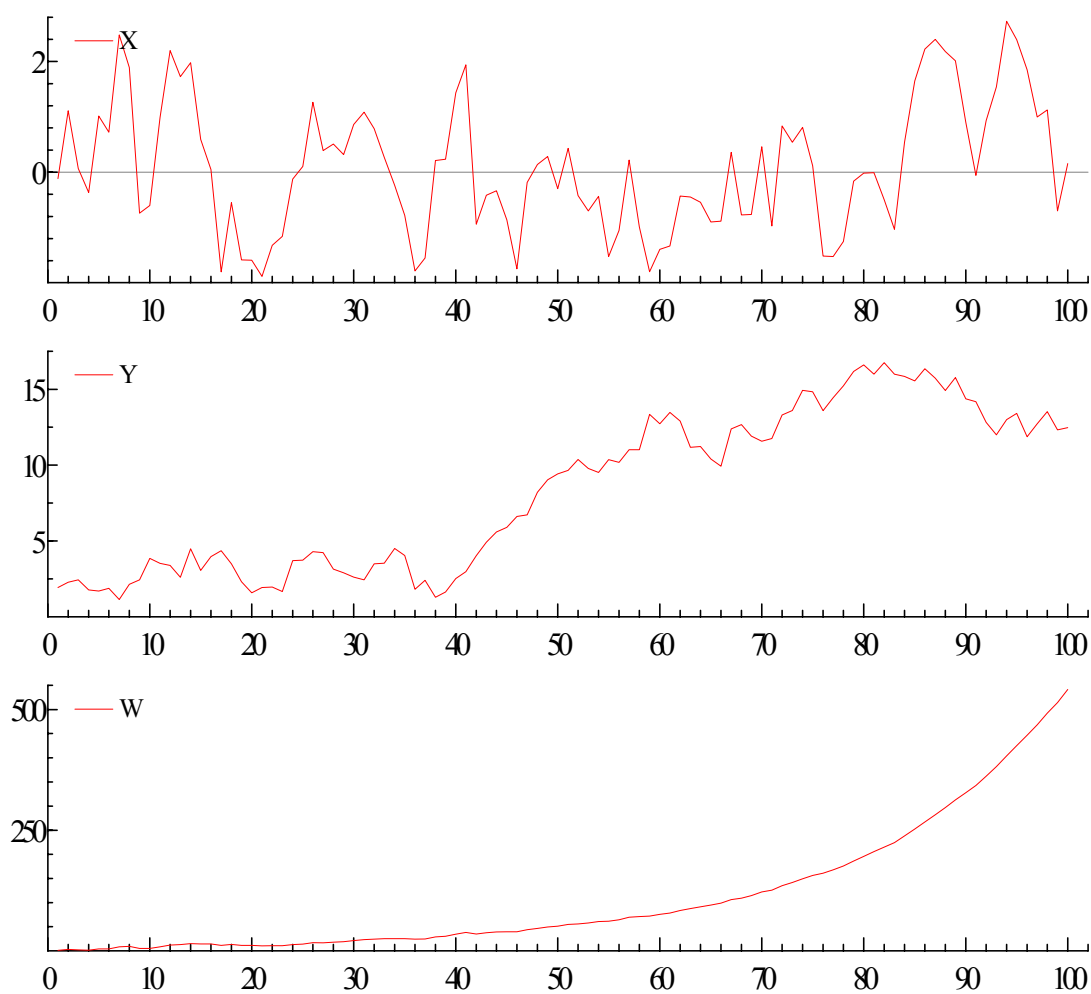


Figure 1: Exemplos de séries geradas por modelos AR(1)

Todavia, como sob H_0 y_t é não estacionária, a distribuição desta estatística não é a usual t -Student nem, assintoticamente, a normal. Mais adiante estudaremos a resolução do problema.

4 Modelos ARMA para Séries Estacionárias

O estudo das propriedades dos diversos tipos de modelos ARMA é importante para seleccionar um (ou alguns) modelo(s) que possam ser considerados adequados para aproximar o comportamento da série temporal em estudo. Ou seja, é o conhecimento dessas propriedades que permite fazer a chamada análise de identificação. Como foram os estatísticos G. E. P. Box e G. M. Jenkins quem, pela primeira vez ⁶, apresentaram um estudo sistemático das referidas propriedades para inúmeros modelos, é usual designar este tipo de modelização como **modelização Box-Jenkins**.

Como se referiu, a análise de identificação é um pouco complexa. Desta forma, o assunto será abordado de forma muito ligeira.

Os modelos autoregressivos são particularmente adequados para situações em que os ϵ_t 's, chamados de **choques ou inovações**, têm um efeito que se vai diluindo no tempo de forma muito lenta. Com efeito, retomando de novo o caso do AR(1) estacionário como exemplo e a equação (1), repare-se que o efeito sobre o nível actual da série de um choque unitário ocorrido há um período atrás é dado por ϕ , o de um choque ocorrido há 2 períodos atrás é de ϕ^2 e, em geral, o de um choque ocorrido há s períodos atrás é de ϕ^s . Assim, se ϕ for elevado (por exemplo, 0.90), pode ser necessário que decorra muito tempo até que a série se "esqueça" do que ocorreu no passado. Refira-se ainda que essa equação é a chamada representação MA(∞) do processo AR(1) e que, no caso não estacionário com $|\phi| > 1$ ter-se-ia um efeito estranho de a série se recordar melhor dos choques mais antigos do que dos mais recentes.

Embora ilustrado apenas com base no caso do AR(1), o **problema da estacionariedade** coloca-se relativamente a todos os processos autoregressivos. Assim, no caso geral de um AR(p),

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t,$$

note-se que escrevendo toda a parte autoregressiva do lado esquerdo da equação tem-se:

$$(1 - \phi_1 L - \dots - \phi_p L^p) y_t = \epsilon_t.$$

Ou seja, fazendo $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$, pode escrever-se

$$\phi(L) y_t = \epsilon_t,$$

⁶No livro *Time Series Analysis, Forecasting and Control*, publicado em 1970.

que é a representação mais comum de um processo autoregressivo.

As condições de estacionaridade podem então ser apresentadas da forma que se segue. Comece por se formar a equação característica:

$$(1 - \phi_1 z - \dots - \phi_p z^p) = 0,$$

isto é, $\phi(z) = 0$. A resolução desta equação fornece as suas raízes, representadas com z_i ($i = 1, 2, \dots, p$), e estas podem ser complexas para os casos $p \geq 2$, isto é, as raízes são do tipo $z_i = a + bi$, com $i^2 = -1$ (isto é, $i = \sqrt{-1}$). Os módulos das raízes são dados por $|z_i| = \sqrt{a^2 + b^2}$. As condições de estacionaridade resumem-se a: sejam reais ou complexas, (todas) as raízes devem ter módulo superior a um, isto é, $|z_i| > 1$ ($i = 1, 2, \dots, p$), dizendo-se que se encontram fora do círculo unitário.

Os modelos autoregressivos serão retomados mais adiante, estudando-se então uma forma muito simples de determinação da sua ordem (isto é, de escolha do p).

Pelo contrário, os processos MA têm memória curta e são sempre estacionários. Por exemplo, para o MA(1) verifica-se facilmente que a série se “esquece” dos choques ocorridos há mais que um período. Para um MA(2), a série “esquece-se” dos choques ocorridos há mais que dois períodos atrás e, em geral, para um MA(q) a série “esquece-se” dos choques ocorridos há mais que q períodos atrás. A estacionaridade de qualquer processo MA finito também é de fácil verificação.

O problema que pode afectar os processos MA é o da **invertibilidade**. Para o ilustrar considere-se o caso do MA(1): $x_t = \epsilon_t - \theta \epsilon_{t-1}$, isto é, $x_t = (1 - \theta L) \epsilon_t$. Resolvendo em ordem a ϵ_t :

$$\epsilon_t = (1 - \theta L)^{-1} x_t,$$

isto é, desde que $|\theta| < 1$,

$$\begin{aligned} \epsilon_t &= (1 + \theta L + \theta^2 L^2 + \dots) x_t, \\ &= x_t + \theta x_{t-1} + \theta^2 x_{t-2} + \dots, \end{aligned} \tag{4}$$

que também se pode escrever como

$$x_t = -\theta x_{t-1} - \theta^2 x_{t-2} + \dots + \epsilon_t,$$

isto é, como um processo AR(∞). Note-se que esta representação também só faz sentido se $|\theta| < 1$. No caso contrário, x_t dependeria mais fortemente dos seus valores mais antigos do que dos mais recentes ($|\theta| > 1$) ou dependeria da mesma forma de todos os valores passados ($|\theta| = 1$). Diz-se então que se inverteu o processo MA(1) e a

condição $|\theta| < 1$ é chamada de condição de invertibilidade do MA(1). Refira-se também que, se esta condição não for satisfeita, poderão ocorrer problemas de estimação e de previsão.

Mais geralmente, considere-se o processo MA(q), $x_t = \epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_q\epsilon_{t-q}$, que se pode escrever como

$$x_t = (1 - \theta_1 L - \dots - \theta_q L^q)\epsilon_t,$$

isto é,

$$x_t = \theta(L)\epsilon_t, \text{ com } \theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q.$$

Procedendo como anteriormente, $\theta(L)^{-1}x_t = \epsilon_t$, com $\theta(L)^{-1}$ um polinómio infinito. Resolvendo em ordem a x_t obtém-se a representação AR(∞) do processo MA, do tipo,

$$x_t = \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \dots + \epsilon_t,$$

que só tem sentido se todas as raízes do polinómio $\theta(L)$, isto é, se todas as soluções da equação característica $\theta(z) = 0$, se situarem fora do círculo unitário. Estas são as condições de invertibilidade do processo MA. Tal como para o MA(1), se estas condições não forem satisfeitas, podem observar-se problemas de estimação e de previsão.

Finalmente, para o caso mais geral dos processos ARMA (estacionários), tem-se

$$\phi(L)y_t = \theta(L)\epsilon_t,$$

com $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ o polinómio autoregressivo de ordem p e $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ o polinómio MA de ordem q . Satisfeitas as condições de estacionaridade (raízes de $\phi(L)$ fora do círculo unitário), o processo pode ser escrito como um MA(∞),

$$y_t = \phi(L)^{-1}\theta(L)\epsilon_t,$$

com coeficientes que tendem para zero à medida que vamos recuando no tempo.

Como exemplo, considere-se o processo ARMA(1,1) dado por $y_t = 2 + 0.8y_{t-1} + \epsilon_t - 0.5\epsilon_{t-1}$, isto é,

$$(1 - 0.8L)y_t = 2 + (1 - 0.5L)\epsilon_t.$$

Resolvendo em ordem a y_t , invertendo o seu polinómio autoregressivo,

$$y_t = (1 - 0.8L)^{-1}2 + (1 - 0.8L)^{-1}(1 - 0.5L)\epsilon_t,$$

isto é,

$$y_t = 2/(1 - 0.8) + (1 - 0.5L)(1 + 0.8L + 0.8^2 L^2 + \dots)\epsilon_t,$$

ou seja,

$$y_t = 10 + (1 + 0.8L + 0.8^2L^2 + 0.8^3L^3 + \dots - 0.5L - 0.4L^2 - 0.32L^3 - \dots)\epsilon_t.$$

Desta forma, passado um período, o efeito de um choque unitário é de 0.3 unidades, passados dois períodos é de 0.24 unidades, passados três períodos é de 0.192, etc..

Por outro lado, satisfeitas também as condições de invertibilidade (raízes de $\theta(L)$ fora do círculo unitário), o processo também pode ser escrito como um AR(∞),

$$\theta(L)^{-1}\phi(L)y_t = \epsilon_t,$$

também com coeficientes que vão tendendo para zero à medida que nos afastamos do presente. Desta forma, a partir de certa ordem, os coeficientes serão tão pequenos que, para efeitos práticos, poderão ser negligenciados. Ou seja, mesmo um processo misto ARMA poderá ser aproximado por um AR finito. É esta aproximação que justifica que, mais adiante, apenas nos preocupemos com os processos (modelos) AR finitos.

5 Previsão

Como se referiu inicialmente, o principal objectivo da utilização dos modelos ARMA é a projecção ou extrapolação da série para o futuro, para além do período da amostra.

Obviamente, a previsão está sujeita a erros. Os erros de previsão têm duas fontes, ou origens:

1. o desconhecimento dos choques ou inovações futuras;
2. o desconhecimento dos valores dos parâmetros e os erros inevitáveis da sua estimação.

Como em JD, no que se segue só consideraremos a primeira fonte. Também só alguns modelos simples serão estudados a título de exemplo.

Assume-se que temos observações para a variável de interesse (y) de 1 a n . Assim, as previsões são efectuadas de forma **condicional na informação disponível** no final do período n , isto é, o final da amostra é a origem da previsão.

Considerem-se as seguintes notações:

- y_{n+s} representa o valor desconhecido (a prever) no período $n + s$;
- assim, s representa o horizonte de previsão;

- \hat{y}_{n+s} representa a previsão de y_{n+s} feita com base na informação disponível no final do período n (mas mais adiante representará a previsão ótima).

Suponha-se que se pretende que as previsões obedeçam às seguintes condições:

- a) os erros de previsão positivos e os negativos implicam custos idênticos;
- b) todavia, pretende-se evitar grandes erros de previsão, penalizando-os fortemente.

Nestas condições, para avaliar a qualidade das previsões é razoável usar o seu erro quadrático médio,

$$EQM(\hat{y}_{n+s}) = E[(y_{n+s} - \hat{y}_{n+s})^2],$$

e, naturalmente, procura-se encontrar a previsão que minimize o EQM (de previsão). Pode então mostrar-se que **a previsão de y_{n+s} , efectuada no final do período n , que minimiza o EQM, é o valor esperado de y_{n+s} condicional em toda a informação disponível até esse momento.** Por outras palavras, a previsão ótima (em EQM) é dada por

$$\hat{y}_{n+s} = E_n(y_{n+s}),$$

onde o índice n em E traduz o condicionamento sobre a informação disponível até ao referido momento.⁷

Como primeiro exemplo de aplicação desta regra, comecemos por considerar o caso do processo **AR(1)**, $y_t = m + \phi y_{t-1} + \epsilon_t$. Mesmo que s seja elevado, como a previsão é efectuada recursivamente, teremos que obter inicialmente a previsão para $s = 1$, depois para $s = 2$, etc.. Então, para $s = 1$ tem-se

$$y_{n+1} = m + \phi y_n + \epsilon_{n+1}.$$

Logo,

$$\hat{y}_{n+1} = E_n[m + \phi y_n + \epsilon_{n+1}].$$

Como calcular o valor esperado condicional? Relativamente ao presente e ao passado, isto é, no final do período n , os valores das variáveis aleatórias referentes a esses períodos foram já observados, isto é, são (considerados como) conhecidos. Assim,

$$\text{para } t \leq n, \quad E_n(y_t) = y_t \quad \text{e} \quad E_n(\epsilon_t) = \epsilon_t.$$

⁷ Outra forma de representar o mesmo, muito comum na literatura, é $E(y_{n+s}|I_n)$, onde I_n representa toda a informação disponível até ao final do período n . Recorde-se, contudo, que a informação aqui empregue se refere apenas à história da variável e das inovações ou choques.

Mas relativamente aos ϵ_t 's futuros ($t > n$), o valor esperado condicional é igual ao valor esperado não condicional (zero) dado que a independência dos ϵ 's implica que o conhecimento do seu passado (e presente) nada adianta sobre o futuro:

$$\text{para } t > n, \quad E_n(\epsilon_t) = E(\epsilon_t) = 0.$$

Desta forma, a previsão óptima para y_{n+1} é

$$\hat{y}_{n+1} = E_n[m + \phi y_n + \epsilon_{n+1}] = m + \phi y_n.$$

Por conseguinte, o erro de previsão é ϵ_{n+1} . Não admira, assim, que os ϵ 's sejam designados como inovações, ou "surpresas", ou choques.

Por outro lado, para $s = 2$, tem-se

$$y_{n+2} = m + \phi y_{n+1} + \epsilon_{n+2}.$$

Por conseguinte,

$$\begin{aligned} \hat{y}_{n+2} &= E_n[m + \phi y_{n+1} + \epsilon_{n+2}] \\ &= m + \phi \hat{y}_{n+1} \\ &= m + \phi(m + \phi y_n) \\ &= m + \phi m + \phi^2 y_n. \end{aligned}$$

Em geral, para um s (positivo) qualquer, tem-se

$$\hat{y}_{n+s} = m + \phi \hat{y}_{n+s-1},$$

que também se pode mostrar ser

$$\hat{y}_{n+s} = \mu + \phi^s (y_n - \mu),$$

com $\mu = m/(1 - \phi)$ (reveja-se a equação (2)). Esta expressão permite concluir que, à medida que o horizonte de previsão aumenta, como $|\phi| < 1$ a previsão tende para a média do processo. Ou seja, as previsões para prazos um pouco mais longos tornam-se pouco informativas. Esta é uma das razões que levam a desaconselhar a utilização destes modelos para previsões de médio e de longo prazos.

Ainda a título de exemplo considere-se agora o caso do processo **MA(1)**:

$$y_t = m + \epsilon_t - \theta \epsilon_{t-1},$$

para o qual se verifica facilmente que $\mu = E(y_t) = m$. Então, para $s = 1$:

$$\hat{y}_{n+1} = E_n[m + \epsilon_{n+1} - \theta\epsilon_n] = m - \theta\epsilon_n.$$

Para $s = 2$,

$$\hat{y}_{n+2} = E_n[m + \epsilon_{n+2} - \theta\epsilon_{n+1}] = m = \mu.$$

E também se verifica facilmente que, para $s > 2$, também se tem $\hat{y}_{n+s} = m = \mu$. Ou seja, as previsões tornam-se pouco informativas muito rapidamente (e daí que os modelos MA puros sejam pouco utilizados para as séries económicas).

Finalmente, refira-se que:

1. A incerteza de previsão é medida através da variância do erro de previsão. Ora, pode mostrar-se que:
 - a) para processos ARMA (estacionários), a variância dos erros de previsão é limitada e tende para a variância não condicional do processo.
 - b) Pelo contrário, para os processos não estacionários a variância dos erros de previsão não é limitada e vai crescendo com o horizonte de previsão. Este comportamento parece muito mais plausível para muitas das variáveis económicas: a incerteza de previsão aumenta à medida que nos afastamos do presente. Isto é, muita atenção deve ser dada aos processos não estacionários.
2. Em termos práticos, é evidente que os parâmetros não são conhecidos e necessitam ser substituídos por estimativas nas expressões anteriores. Também os ϵ_t 's, com $t \leq n$, não são efectivamente observados. Assim, na prática são substituídos pelos $\hat{\epsilon}_t$'s (ou e_t 's), isto é, pelos resíduos ou erros de estimação.

6 Modelos Autoregressivos

Pelas razões já referidas e também porque são mais fáceis de estimar, apenas os modelos autoregressivos serão objecto de atenção mais detalhada. Não se pense, no entanto, que os modelos MA e ARMA não são úteis para as séries económicas. Por exemplo, um modelo que parece adequar-se bem a algumas séries económicas é

$$\Delta y_t = m + \epsilon_t - \theta\epsilon_{t-1},$$

com $\theta > 0$, isto é, um modelo MA(1) sobre a série diferenciada (isto é, um modelo ARIMA(0,1,1)) ⁸.

6.1 Estimação

Considere-se o modelo AR(p):

$$y_t = m + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t.$$

Como os erros do modelo seguem um processo ruído branco, o estimador OLS é consistente. Adicionalmente, desde que a distribuição dos ϵ_t 's seja normal ⁹, pode mostrar-se que o estimador OLS é idêntico ao estimador de máxima verosimilhança condicional nas primeiras p observações. Desta forma, o estimador OLS pode também ser considerado como assintoticamente eficiente.

6.2 Selecção

Os resultados anteriores assumem que a ordem do processo, p , é conhecida. Todavia, um dos problemas mais importantes para os modelos autoregressivos é precisamente o da escolha (ou estimação) da sua ordem. Contudo, não parece existir ainda uma solução óptima, unanimemente aceite.

Assim, muitos investigadores usam a estatística AIC, escolhendo o modelo que minimiza essa estatística. Outros empregam, por exemplo, a estatística SC (ou SIC ou SBIC) para o mesmo efeito. Outras estatísticas do mesmo tipo também são empregues. Note-se, contudo, que o R^2 não deve ser empregue. De facto, por exemplo, para o processo AR(1), $y_t = m + \phi y_{t-1} + \epsilon_t$, o R^2 da população é dado por ϕ^2 . Ou seja, se, por exemplo, $\phi = 0.9$, o R^2 é muito maior do que se $\phi = 0.2$, mas esse facto não significa que o primeiro modelo seja mais adequado que o segundo. Por outro lado, também não se deve comparar, usando o R^2 , a qualidade de um modelo em que a variável dependente aparece diferenciada com a de outro em que aparece sem o ser (em nível). Porquê?

Nos últimos anos, a chamada **estratégia de modelização do geral para o particular** tem vindo a prevalecer no trabalho empírico. Frequentemente, esta estratégia serve-se de testes- t individuais, de significância sobre os coeficientes autoregressivos, sendo por isso chamada também de “ t -sig”.

⁸O “I” que agora aparece e o correspondente “1” referem-se ao facto de a série ser integrada de primeira ordem.

⁹Dizendo-se que $\{\epsilon_t\}$ é um processo ruído branco Gaussiano.

Em termos simples, esta estratégia parte da fixação de uma ordem máxima para a autoregressão, seja p_{max} . Estimado o modelo autoregressivo com essa ordem, testa-se a significância do último coeficiente, isto é, do coeficiente $\phi_{p_{max}}$, utilizando um vulgar rácio- t , válido assintoticamente. Se a hipótese nula for rejeitada o processo pára imediatamente e é seleccionado $p = p_{max}$. No caso contrário, o modelo é simplificado com a exclusão do último desfasamento e re-estimado. O processo de teste é então retomado: testa-se a significância do último coeficiente ($\phi_{p_{max}-1}$) e, caso a hipótese nula seja rejeitada, o processo pára. No caso contrário, o modelo é reespecificado com a exclusão do último desfasamento e o procedimento de estimação e de teste retomado. Ou seja, o processo sequencial só pára quando se obtém a rejeição de uma hipótese nula (ou quando se atinge a ordem zero).

Daqui emergem dois novos problemas: a) que dimensão usar para os testes? b) como escolher p_{max} ? Ambos os problemas têm sido resolvidas de forma algo arbitrária. Relativamente ao primeiro, muitos investigadores usam o habitual $\alpha = 0.05$ mas outros preferem usar $\alpha = 0.10$. Como é evidente, com alguma frequência os últimos tenderão a seleccionar ordens mais elevadas que os primeiros. Ou seja, quanto mais elevado for o α empregue nos testes sequenciais mais "liberal" tenderá a ser a especificação escolhida.

Por outro lado, para escolher p_{max} várias regras determinísticas têm sido sugeridas, todas elas associando essa ordem à dimensão da amostra. Uma regra algo popular é a chamada l_{12} , que propõe que

$$p_{max}(n) = [12(n/100)^{1/4}]$$

onde $[.]$ representa "a parte inteira de". Por exemplo, supondo que $n = 50$ ter-se-á $p_{max} = 10$. Todavia, outros investigadores preferem usar regras que não implicam a perda de tantas observações (como a regra " l_4 "). Assim, por exemplo, é frequente usar-se $p_{max} = 6$ quando $n = 50$ e $p_{max} = 10$ quando $n = 100$.¹⁰

6.3 Testes-diagnóstico

O modelo escolhido e estimado com base nos procedimentos referidos deverá ainda ser submetido a "testes-diagnóstico" e, em particular, a testes de autocorrelação. A ideia base é de que se os resíduos do modelo apresentam ainda alguma autocorrelação é

¹⁰Note-se que a utilização do procedimento sequencial t -sig tem dominado a prática econométrica nos últimos 10 anos mas que se descobriu recentemente que o método funciona mal quando o modelo autoregressivo contém regressores determinísticos, e é tanto pior quanto maior for o número destes. Assim, dentro de poucos anos, é provável que este procedimento seja abandonado.

porque a estrutura dinâmica do modelo é inadequada ou insuficiente. De outra forma, num modelo especificado correctamente, os erros deverão evoluir segundo um processo ruído branco. Assim, analisam-se os resíduos de estimação, que deverão reflectir (aproximadamente) este comportamento ideal.

Embora na análise tradicional das séries temporais sejam empregues mais frequentemente os testes de Ljung-Box, são os testes de Breusch-Godfrey (de multiplicadores de Lagrange) que deverão ser usados preferencialmente. De facto, alguns estudos têm demonstrado que os últimos são mais potentes que os primeiros. Adicionalmente, relembre-se que convém efectuar testes para várias ordens: o teste de primeira ordem é obrigatório mas (pelo menos) um teste de ordem superior — testando a ausência de autocorrelação até à ordem 3 ou 4 —, também deve ser efectuado. Adicionalmente, a frequência das observações também dá indicações úteis. Por exemplo, para dados trimestrais é praticamente obrigatório um teste para as autocorrelações até à ordem 4 (ou até 8, se o número de observações for grande); para observações mensais devem considerar-se testes até às ordens 6 e 12, etc..

6.4 Exemplo

Para exemplo simples de modelização e previsão autoregressiva consideraram-se os dados trimestrais (do logaritmo) do consumo de cimento no mercado interno, do exercício 40, entre 80.1 e 2001.4, cujo gráfico é apresentado na “figure 2”. Da análise gráfica é claro que se deve incluir nos modelos um termo de tendência determinística bem como um conjunto de “dummies” sazonais. Por outro lado, note-se que o trabalho de especificação se baseou apenas nos dados de 80.1 a 2001.2, reservando para a simulação das previsões as duas últimas observações.

Como primeiro modelo (M1) considerou-se uma autoregressão com $p_{max} = 8$, tendo-se obtido $p = 6$ (ou $\hat{p} = 6$) pelo processo sequencial, t -sig, de eliminação dos desfazamentos insignificantes. O modelo estimado é

$$y_t = 1.032 + 0.070Q_{1t} + 0.107Q_{2t} + 0.084Q_{3t} + 0.002t + 0.549y_{t-1} + 0.260y_{t-2} - 0.025y_{t-3} + 0.138y_{t-4} + 0.175y_{t-5} - 0.252y_{t-6} + \hat{u}_{1t}, \hat{\sigma}_{u_1} = 0.052,$$

mas com os regressores y_{t-3} , y_{t-4} e y_{t-5} algo longe da significância estatística (os valores- p dos seus coeficientes são, respectivamente, 0.857, 0.319 e 0.202). Este modelo ajusta-se bem aos dados ($R^2 = 0.96$) e, sobretudo, “passa” facilmente os testes BG de autocorrelação de primeira ordem e até à ordem quatro: $BG(1) = 0.74$ [0.388] e $BG(4) = 3.61$ [0.459], com os valores dentro de parêntesis a representarem os valores- p .

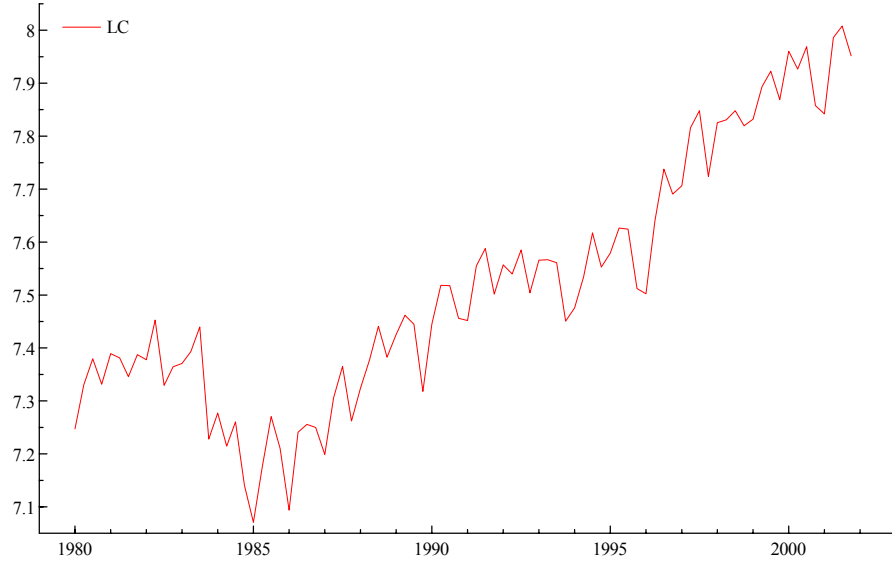


Figure 2: Logaritmo do consumo de cimento no mercado interno

Como segundo modelo (M2) considerou-se uma autoregressão com $p_{max} = 4$, que produziu

$$y_t = 0.667 + 0.059Q_{1t} + 0.120Q_{2t} + 0.106Q_{3t} + 0.001t + 0.620y_{t-1} + 0.276y_{t-2} + \hat{u}_{2t}, \quad \hat{\sigma}_{u_2} = 0.053,$$

ou seja, um modelo bastante mais parcimonioso, apenas com $p = 2$, o que ilustra o facto bem conhecido de a ordem seleccionada do processo autoregressivo depender fortemente da ordem máxima inicialmente assumida. Este modelo também se ajusta bem aos dados ($R^2 = 0.95$) e também “passa” facilmente nos testes de autocorrelação: $BG(1)=0.05[0.825]$ e $BG(4) = 5.44[0.245]$.

Finalmente, como nos modelos anteriores as somas dos coeficientes autoregressivos não se encontram muito longe da unidade, especificou-se e estimou-se um modelo nas primeiras diferenças (isto é, impondo uma raiz unitária), iniciando o processo com três desfasamentos. (Porque não também quatro desfasamentos, como no caso anterior?) O modelo a que se chegou é simplesmente

$$\Delta y_t = -0.064 + 0.063Q_{1t} + 0.127Q_{2t} + 0.109Q_{3t} - 0.320\Delta y_{t-1} + \hat{u}_{3t}, \quad \hat{\sigma}_{u_3} = 0.053,$$

com todos os coeficientes “altamente” significativos ¹¹. O $R^2 = 0.494$ pode ser consi-

¹¹Porque não se considerou o termo de tendência determinística neste modelo?

derado como bastante elevado atendendo a que se refere a uma série diferenciada. Por outro lado, o modelo também passa facilmente os testes de autocorrelação: $BG(1) = 0.20[0.657]$ e $BG(4) = 4.73 [0.316]$.

Os três modelos foram empregues para efectuar previsões para os dois últimos trimestres de 2001 ¹², com os resultados que se apresentam no quadro 1. RQEQM e EAM representam, respectivamente, a raiz quadrada do EQM e o erro absoluto médio.

Quadro 1 – Valores observados e previstos (em milhares de toneladas) e análise sumária das previsões

	2001.3	2001.4	RQEQM	EAM
val. obs.	3005	2840	—	—
prev. M1	2874	2717	127.1	127.1
prev. M2	2941	2757	73.8	73.1
prev. M3	2934	2752	79.7	79.3

Desta forma, como se constata facilmente, a ordenação dos modelos em termos de qualidade das previsões é dada por: M2, M3 e M1. Note-se ainda que é possível obter previsões claramente melhores que estas mas com um custo (de obtenção de informação adicional) superior. Aliás, mesmo empregando apenas uma análise univariada um pouco mais sofisticada, também é possível obter previsões de melhor qualidade. Em particular, basta incorporar no modelo a quebra na tendência da série ocorrida ainda na primeira metade da década de 80, que é ainda mais perceptível quando se emprega uma amostra de dados anuais iniciada em 1960 ou mesmo em 1948.

7 Análise de Estacionaridade

Usando como exemplo o caso do AR(1), já vimos que a distinção entre série (processo) estacionária e não estacionária mas integrada é crucial sob vários pontos de vista: momentos (até à ordem dois) e comportamento da série. Também os métodos ¹³ e a incerteza de previsão dependem fortemente da natureza da série.

¹²Recorde que, após a instrução de estimação de cada modelo pelo OLS, é bastante fácil obter as respectivas previsões usando o TSP. Por exemplo, neste caso: "SMPL 2001:3 2001:4; FORCST PREV; PRINT PREV;". Por outro lado, inúmeras estatísticas de análise da qualidade das previsões podem ser obtidas com o comando ACTFIT. Por exemplo: "ACTFIT OBS PREV;", com "OBS" e "PREV" a representarem, respectivamente, as séries de observações e de previsões.

¹³Este ponto não foi desenvolvido mas o seu tratamento é bastante simples. Suponha-se, por exemplo, que y_t é I(1), do tipo $y_t = m + y_{t-1} + u_t$, com u_t um processo ARMA (estacionário). Então,

Por outro lado, também o efeito de longo prazo dos choques depende dessa natureza. Usando como exemplo de novo o caso do AR(1) estacionário ($|\phi| < 1$) e recuperando a equação (1),

$$y_t = m(1 + \phi + \phi^2 + \phi^3 + \dots) + (\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \dots),$$

repare-se que os choques têm um efeito **transitório**, que se vai diluindo com a passagem do tempo. Em termos formais: $\partial y_t / \partial \epsilon_{t-s} = \phi^s$. Por conseguinte, $\lim_{s \rightarrow \infty} \partial y_t / \partial \epsilon_{t-s} = \lim_{s \rightarrow \infty} \phi^s = 0$. Pelo contrário, para um processo integrado de primeira ordem, o efeito dos choques ou inovações é **permanente**, não se vai diluindo com a passagem do tempo¹⁴. Usando de novo o caso do passeio aleatório para ilustração, relembre-se a equação (3):

$$y_t = y_0 + mt + \sum_{i=1}^t \epsilon_i.$$

Ou seja, supondo que o processo foi iniciado há muito tempo, $\lim_{s \rightarrow \infty} \partial y_t / \partial \epsilon_{t-s} = 1$. É como se a economia se “lembrasse” hoje completamente de todos os choques, mesmo daqueles que ocorreram há muito tempo no passado.

Por outro lado, ainda, os métodos de estimação e de inferência tradicionais, formulados para séries estacionárias, em geral funcionam mal para as séries não estacionárias. No capítulo 5, a propósito da análise multivariada, estudar-se-á uma motivação adicional.

7.1 Séries económicas: estacionárias em tendência (TSP) ou por diferenciação (DSP)?

Embora útil e motivadora, a análise anterior não é a mais relevante para séries económicas como as do produto, do consumo, do investimento, do índice de preços, do rendimento, da moeda, etc.. Todas estas séries têm uma característica comum: uma tendência (ou comportamento de longo prazo) crescente com o tempo. Então, obviamente, como não têm média constante ao longo do tempo, não são estacionárias.

Todavia, dois modelos muito distintos estão disponíveis para aproximar este tipo de comportamento: o modelo (ou processo) de estacionaridade em tendência (**TSP**,

a previsão para y_{t+1} é dada por $\hat{y}_{t+1} = m + y_t + \hat{u}_{t+1}$, com $\hat{u}_{t+1} = E_t(u_{t+1})$.

¹⁴Refira-se que este aspecto não parece ser completamente razoável: embora a economia pareça ser por vezes afectada por choques permanentes, alguns (muitos) choques têm natureza transitória, não afectando de forma duradoura.

trend stationary process) e o de estacionaridade por diferenciação (**DSP**, *difference stationary process*) ou de raiz unitária (**RU**).

Segundo a hipótese TSP, a série y_t pode escrever-se como

$$y_t = f(t) + u_t,$$

onde $f(t)$ é uma função do tempo e u_t representa um processo (ARMA) estacionário. Por exemplo, assumindo uma tendência linear determinística, que é a hipótese tradicionalmente considerada, ter-se-ia

$$y_t = \beta_0 + \beta_1 t + u_t.$$

Todavia, como as séries económicas têm, em geral, uma taxa de crescimento aproximadamente constante, o modelo de tendência exponencial revela-se, frequentemente, mais adequado:

$$y_t = \exp[\beta_0 + \beta_1 t + u_t],$$

donde resulta

$$\ln(y_t) = \beta_0 + \beta_1 t + u_t, \quad (5)$$

isto é, com a presença da tendência linear na série logaritmizada.¹⁵ Em ambos os casos, o aspecto a realçar é que os desvios em relação à tendência (u_t) são assumidos como estacionários. Ora esta visão corresponde à perspectiva tradicional dos ciclos económicos: a tendência determinística seria determinada pelo crescimento quantitativo e qualitativo da população, pela acumulação de capital, pelo progresso tecnológico, etc.. Os ciclos económicos representariam apenas perturbações transitórias relativamente a essa tendência. Por outras palavras, os efeitos dos choques económicos dissipar-se-iam mais ou menos rapidamente com o tempo.

Pelo contrário, teorias mais recentes como a dos ciclos económicos reais — que atribuem a totalidade das flutuações do produto aos choques de produtividade —, ou do rendimento permanente, por exemplo, salientam a forte persistência dos choques sobre variáveis como o produto e o consumo, respectivamente. Assim, a melhor maneira de representar estatisticamente essas séries seria através de processos integrados, isto é, estacionários por diferenciação (DSP) ou de raiz unitária (RU).

De facto, retomando, por exemplo, o caso do passeio aleatório com deriva, $y_t = \beta_1 + y_{t-1} + \epsilon_t$ e resolvendo recursivamente, já vimos que a presença da tendência crescente (desde que $\beta_1 > 0$) também é facilmente acomodada por um processo integrado

¹⁵No que se segue, utilizar-se-á apenas y_t mesmo que se trate de um série previamente logaritmizada.

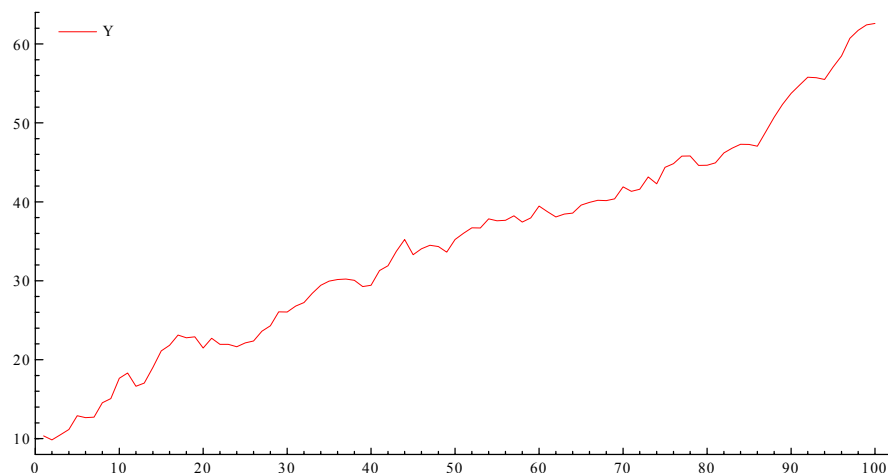


Figure 3: Exemplo de série estacionária por diferenciação

ou de raiz unitária: $y_t = y_0 + \beta_1 t + \sum_{i=1}^t \epsilon_i$. Como esta equação mostra, a grande diferença relativamente ao modelo de estacionaridade em tendência (TS) é que a presença da RU introduz também uma tendência estocástica. Ou seja, os desvios relativamente à tendência determinística são, segundo a hipótese DSP ou de raiz unitária, não estacionários. Na “figure” 3 apresenta-se um exemplo de uma série gerada com base num passeio aleatório com deriva positiva, cujo aspecto é semelhante ao de muitas séries macroeconómicas.

A questão a responder é, pois, a seguinte: os desvios em relação à tendência determinística são meramente transitórios, existindo como que uma “mão invisível” que volta a colocar a economia no caminho da tendência determinística ou, pelo contrário, os choques económicos têm um efeito permanente?

Na fase inicial da “revolução das raízes unitárias” (nas década de 80 e em boa parte da de 90), foram sobretudo estas motivações decorrentes do debate (político-)económico que a alimentaram. Todavia, como já vimos, os dois modelos têm também implicações estatísticas claramente dissemelhantes. Por outro lado, como é óbvio, o procedimento adequado para a estacionarização depende da natureza da série: a) se esta for TS deve remover-se a tendência determinística, usando-se os resíduos OLS da regressão (4); b) no caso DSP, obviamente, deve diferenciar-se a série.

Saliente-se que é incorrecto estacionarizar uma série TS diferenciando-a e estacionarizar uma série DS removendo a tendência determinística:

- relativamente ao primeiro caso, por exemplo se $y_t = \beta_0 + \beta_1 t + \epsilon_t$, se diferenciarmos

a série obtém-se $\Delta y_t = \beta_1 + \epsilon_t - \epsilon_{t-1}$, isto é, a diferenciação introduz uma raiz unitária no polinómio MA, que é não invertível;

- relativamente à segunda afirmação, e tomando como exemplo o caso do passeio aleatório com deriva, $y_t = m + y_{t-1} + \epsilon_t$, já vimos que ele se pode escrever como $y_t = y_0 + mt + \sum_{i=1}^t \epsilon_i$ (veja-se a equação (3)). Desta forma, como a tendência estocástica se mantém, a remoção da tendência determinística é insuficiente para estacionarizar o processo.

Por outro lado, salienta-se que a análise de estacionaridade deve ser vista preferencialmente como a dos desvios em relação a uma componente determinística: esta pode não existir, ou pode ser dada apenas por uma constante ou por “dummies” sazonais, ou por uma tendência determinística adicionada às anteriores, etc. . Por exemplo, para saber a ordem de integração da série dada por

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + u_t,$$

a presença do polinómio em t é irrelevante: é a natureza de u_t que determina a ordem de integração de y_t .

Assim, a definição de ordem de integração dada anteriormente é agora tornada mais precisa ¹⁶: um processo estocástico, y_t , **sem componentes determinísticos**, é integrado de ordem d , $y_t \sim I(d)$, quando só após ser diferenciado d vezes se torna estacionário (não o sendo ainda após $d - 1$ diferenciações). Desta forma, emprega-se a notação $y_t \sim I(0)$ para representar uma série (processo) que não necessita ser diferenciada nenhuma vez para se tornar estacionária (embora possa ser não estacionária, como no caso TS).

Mais geralmente se, excluindo os termos determinísticos, o processo y_t se escrever como

$$\phi(L)y_t = \theta(L)\epsilon_t,$$

onde $\phi(L)$ e $\theta(L)$ representam, respectivamente, os polinómios autoregressivo e de média móveis, e se o primeiro tiver d raízes unitárias, com as restantes fora do círculo unitário, então pode escrever-se

$$(1 - L)^d \gamma(L)y_t = \theta(L)\epsilon_t,$$

onde todas as raízes de $\gamma(L)$ se encontram fora do círculo unitário, isto é,

$$\gamma(L)\Delta^d y_t = \theta(L)\epsilon_t.$$

¹⁶Segundo Engle, R. F. e Granger, C. W. J. (1987), Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica*, 55, 251-76.

Assim, determinar a ordem de integração de uma série é o mesmo que determinar o número de raízes unitárias do seu polinómio autoregressivo, e daí a designação de raízes unitárias autoregressivas e de testes de raízes unitárias (autoregressivas). Em geral, há evidência empírica de que muitas séries económicas são bem aproximadas por processos $I(1)$; algumas “serão $I(0)$ ” e muito poucas, apenas algumas séries nominais como as dos índices de preços ou do *stock* de moeda, aparentam ser $I(2)$. Na análise que se segue, só as hipóteses $I(1)$ e $I(0)$ serão consideradas.

7.2 Testes de raízes unitárias: Testes DF e ADF

Pelas razões apontadas e também pela sua importância como instrumento preliminar da análise multivariada, os testes de raízes unitárias tornaram-se muito importantes nos últimos 20 anos. Inúmeros testes foram propostos mas ainda hoje são os testes de Dickey e Fuller os mais importantes nas aplicações empíricas.

7.2.1 Testes DF

Inicia-se a análise com o teste mais comum, para escolher entre o modelo TS

$$y_t = \beta_0 + \beta_1 t + u_t,$$

com u_t um processo AR(1) estacionário, e o modelo DS mais simples com deriva, isto é, com uma RU:

$$y_t = \beta_1 + y_{t-1} + \epsilon_t.$$

Como os dois modelos são não encaixados, é necessário especificar um modelo mais geral de que ambos resultem como caso particular. Pode mostrar-se que esse modelo mais geral é do tipo

$$y_t = \alpha + \beta t + \rho y_{t-1} + \epsilon_t, \quad (6)$$

com $\epsilon_t \sim \text{iid}$, obtendo-se o TSP quando $\rho < 1$ e o DSP quando $\rho = 1$.

Trata-se então de testar a hipótese nula de uma RU com base nesta equação, isto é,

$$H_0 : \rho = 1 \Leftrightarrow y_t \sim I(1),$$

contra a alternativa de estacionaridade em tendência,

$$H_1 : \rho < 1 \Leftrightarrow y_t \sim I(0),$$

e note-se que o teste é **unilateral esquerdo**: a região crítica será toda definida na aba esquerda da distribuição da estatística de teste. Esta é uma estatística- t , representada com $\tau_{ct} = (\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}}$.

Todavia, a estatística de teste calcula-se ainda mais facilmente se subtrairmos y_{t-1} a ambos os membros de (6):

$$\Delta y_t = \alpha + \beta t + (\rho - 1)y_{t-1} + \epsilon_t,$$

isto é,

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \epsilon_t, \quad (7)$$

onde se fez $\gamma = (\rho - 1)$. Ou seja, trata-se agora simplesmente de testar

$$H_0 : \gamma = 0 \text{ vs } H_1 : \gamma < 0,$$

teste que se pode basear no rácio- t de γ : $\tau_{ct} = \hat{\gamma}/\hat{\sigma}_{\hat{\gamma}}$.

Como mostraram Dickey e Fuller, a distribuição da estatística de teste (sob H_0) não é uma t -Student nem uma normal (nem mesmo assintoticamente) ¹⁷. A distribuição assintótica é algo complicada, envolvendo os chamados processos de Wiener (ou movimentos Brownianos), não simétrica e com mediana e moda menores que zero, e foi primeiramente tabelada por Dickey e Fuller (usando métodos de simulação) ¹⁸. Daí que o nome adoptado para a distribuição tenha sido o desses investigadores, o mesmo acontecendo com os testes, vulgarmente conhecidos como **testes DF**.

Todavia, nem sempre se justifica a inclusão do termo de tendência determinística na equação de teste. É esse o caso, por exemplo, para séries como as das taxas de inflação, ou de desemprego, ou de juro, ¹⁹ etc.. Para casos como esse a questão é: serão os desvios em relação a uma constante não estacionários?

Ou seja, em vez de (7), a regressão de teste é agora

$$\Delta y_t = \alpha + \gamma y_{t-1} + \epsilon_t, \quad (8)$$

continuando a testar-se

$$H_0 : \gamma = 0 \text{ vs } H_1 : \gamma < 0,$$

com base no rácio- t de γ , agora representado com $\tau_c = \hat{\gamma}/\hat{\sigma}_{\hat{\gamma}}$. Note-se, contudo, que também a distribuição da estatística de teste muda, deslocando-se para a direita, isto é,

¹⁷Porque sob H_0 a série é não estacionária, o que invalida a utilização dos métodos de inferência clássicos.

¹⁸Dickey, D. A. e Fuller, W. A. (1979), Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427-31 e Fuller, W. A. (1976), *Introduction to Statistical Time Series*, John Wiley & Sons.

¹⁹Adicionalmente, refira-se que variáveis deste tipo não são logaritmizadas, pois não só não apresentam um comportamento de crescimento exponencial, como já se encontram sob a forma de taxa de variação.

os valores críticos não são os mesmos que para o caso anterior, mas sim mais pequenos em valor absoluto (veja-se o quadro 2).

Finalmente, se a série oscilar em torno de zero (o que raramente ocorre com séries económicas), nem sequer a constante é incluída na equação de teste, que passa a ser dada simplesmente por

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t, \quad (9)$$

representando-se com $\tau_{nc} = \hat{\gamma}/\hat{\sigma}_{\hat{\gamma}}$ a estatística DF. Os valores críticos são ainda mais pequenos em valor absoluto que no caso anterior (veja-se o quadro 2).

No quadro 2 apresentam-se os valores críticos assintóticos mais comuns para os testes de raízes unitárias. Note-se ainda que em alguns trabalhos foram propostos valores críticos exactos (para amostras finitas). Todavia, estes requerem que, além de não autocorrelacionados, os erros das equações de teste tenham distribuição normal e sejam homocedásticos. Desta forma, além de mais simples, a utilização dos valores críticos assintóticos poderá ser mais robusta. Contudo, não se deve perder de vista a sua natureza.

Quadro 2 — Valores críticos assintóticos para os testes DF

estatística	1 %	2.5%	5%	10%
τ_{ct}	-3.96	-3.67	-3.41	-3.13
τ_c	-3.42	-3.12	-2.86	-2.57
τ_{nc}	-2.58	-2.23	-1.95	-1.62

Fonte: Fuller, W. A. (1996), *Introduction to Statistical Time Series*, table 10.A.2, p. 642, John Wiley & Sons.

Na prática, o problema da escolha da regressão de teste a usar ((7), (8) ou (9)) tem implicações mais importantes do que alguns julgam. Assim, por um lado, incluir regressores determinísticos desnecessários (ou irrelevantes) reduz a potência do teste. Isto é, por exemplo, se a regressão adequada for a de (8) mas for empregue a de (7), fornecendo a estatística τ_{ct} , o teste tem menos potência que o baseado na primeira (ou seja, na estatística τ_c). Por outro lado, a exclusão de regressores determinísticos relevantes também reduz a potência dos testes mas, neste caso, de forma ainda mais grave que no anterior, pois esta tende para zero quando $n \rightarrow \infty$ ²⁰.

Como se sugeriu implicitamente, em geral, a escolha entre (7), (8) e (9) pode basear-se numa representação gráfica da série sob análise. Todavia, sobretudo em caso de

²⁰Recorde-se que o comportamento “normal” da potência de um teste contra uma alternativa fixa é de tender para um quando $n \rightarrow \infty$. Assim, neste caso tem-se um comportamento muito mau.

dúvida, o raciocínio económico deve prevalecer sobre a análise gráfica. Assim, por exemplo, uma análise gráfica da série trimestral (ou mensal) da taxa de juro, observada apenas entre os anos 1990 e 1999, sugere a presença de uma tendência determinística decrescente com o tempo. Todavia, essa é apenas uma característica da amostra observada, não fazendo sentido que o modelo adequado para a taxa de juro contenha um termo de tendência determinística. Assim, a regressão de teste a empregar deve ser a de (8) (e não a de (7)). Por outro lado, deve procurar-se trabalhar com amostras que correspondam a um espaço temporal bem maior que os 10 anos deste exemplo.

7.2.2 Testes ADF

Os testes da subsecção anterior poderão, no entanto, não ser válidos. De facto, repare-se que, em todos os casos, o modelo assumido sob a hipótese nula é muito restritivo: trata-se de um simples AR(1), que pode ser insuficiente para capturar a evolução da série em função dos seus valores passados. Por outras palavras, pode ser necessário recorrer a um AR(2) ou, mais geralmente, a um AR(p), com $p > 1$, para conseguir uma aproximação razoável ao comportamento univariado da série. Ora, nesse caso, pode mostrar-se que os erros das equações de teste anteriores serão autocorrelacionados, o que torna a inferência baseada nas estatísticas DF inválida. Por outras palavras, nessas circunstâncias as estatísticas DF não têm (sob H_0) a distribuição de Dickey-Fuller.

A solução proposta por Dickey, Fuller e Said para esse problema consiste em usar uma autoregressão de ordem suficientemente elevada, que permita que os resíduos das equações de teste não apresentem sintomas de autocorrelação. Assim, por exemplo, para o caso DSP *vs.* TSP, em vez de (7) deve empregar-se a equação de teste

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^k \theta_i \Delta y_{t-i} + \epsilon_t, \quad (10)$$

onde k é determinado de forma a que os erros tenham um comportamento próximo do de um processo ruído branco (e continuando a testar-se $H_0 : \gamma = 0$ *vs.* $H_1 : \gamma < 0$)²¹. A justificação para o nome destes testes, “augmented DF”, ou melhor, **ADF**(k), é clara: as equações dos testes DF ((7), (8) ou (9)) são aumentadas com desfasamentos da variável dependente.

Todavia, emerge um novo problema: o da escolha de k . Muito frequentemente, os resultados dos testes serão muito sensíveis a esse valor. Pode também mostrar-se que, com um k mal escolhido, errando-se por defeito, os testes tenderão a ter nível real

²¹Para o caso de um modelo AR(p), pode mostrar-se que $k = p - 1$.

superior ao nominal (usualmente 5% ²²). Por outro lado, um valor para k mais elevado que o estritamente necessário retirará potência aos testes.

Trata-se de uma situação em que a estratégia t -sig, do geral para o particular, tem obtido mais argumentos favoráveis, sobretudo quando o objectivo principal reside no controlo da dimensão dos testes. Assim, inicia-se o processo com um k suficientemente elevado (k_{max}) e vai-se tentando simplificar a autoregressão com testes- t individuais sobre os coeficientes de desfasamento mais elevados, até se obter uma rejeição. Eventualmente poder-se-á chegar a ADF(0), isto é, a uma estatística DF. Todavia, também os testes de Breusch-Godfrey deverão ser usados de forma complementar: se a eliminação de um desfasamento insignificante fizer surgir sintomas de autocorrelação residual até aí inexistentes, esse desfasamento é reintroduzido na equação de teste. Como já se referiu, o valor escolhido para k_{max} depende, em geral, da dimensão da amostra. Assim, por exemplo, para muitas das séries anuais disponíveis para a economia Portuguesa, com $n \approx 40$, poder-se-á iniciar o processo com $k_{max} = 4$ ou 5.

7.3 Exemplo

A título de exemplo vai resolver-se o exercício 43.

No quadro seguinte apresentam-se várias estatísticas obtidas das seguintes equações de regressão, estimadas pelo OLS com base em 40 observações anuais, onde U_t representa a taxa de desemprego de um país da OCDE:

$$\Delta U_t = \alpha + \phi U_{t-1} + \sum_{j=1}^k \gamma_j \Delta U_{t-j} + v_t, \quad (11)$$

$$\Delta U_t = \alpha + \beta t + \phi U_{t-1} + \sum_{j=1}^k \gamma_j \Delta U_{t-j} + w_t. \quad (12)$$

Com base nesses resultados, e justificando devidamente as suas opções, o que pode concluir sobre a ordem de integração de U_t ?

Assim, trata-se de testar $H_0 : U_t \sim I(1)$ vs. $H_1 : U_t \sim I(0)$, ou seja, com a notação apresentada no enunciado,

$$H_0 : \phi = 0 \text{ vs. } H_1 : \phi < 0.$$

E o primeiro problema a resolver é o da escolha da equação: (11) ou (12). Ora, não faz qualquer sentido que uma variável como a taxa de desemprego contenha uma tendência

²²Isto é, tenderão a rejeitar H_0 quando ela é verdadeira mais frequentemente que nos 5% dos casos assumidos.

equa.	k	AIC	$\hat{\phi}$	$\hat{\phi}/\hat{\sigma}_{\hat{\phi}}$	DW	$h - alt$	BG(4)	$\hat{\gamma}_k/\hat{\sigma}_{\hat{\gamma}_k}$
(11)	0	3.42	-1.75	-19.79	1.11	—	33.6[.00]	—
(11)	1	2.98	-0.77	-3.89	2.45	-3.6[.00]	10.9[.03]	-5.43
(11)	2	2.80	-0.66	-3.20	1.74	0.8[.40]	5.8[.21]	-2.96
(11)	3	2.79	-0.55	-2.44	2.00	-1.0[.30]	5.4[.25]	0.85
(11)	4	2.83	-0.60	-2.49	1.94	0.5[.61]	4.0[.41]	1.04
(12)	0	3.45	-1.75	-19.63	1.11	—	29.5[.00]	—
(12)	1	3.01	-0.77	-3.86	2.45	-3.0[.00]	10.1[.04]	-5.39
(12)	2	2.82	-0.67	-3.25	1.75	0.8[.43]	5.7[.22]	-2.97
(12)	3	2.83	-0.57	-2.46	1.98	-1.0[.33]	4.8[.30]	0.79
(12)	4	2.86	-0.62	-2.48	1.94	0.6[.55]	3.8[.43]	0.98

(crescente ou decrescente). Isto é, trata-se de analisar a (não-) estacionaridade em torno de uma constante, e não em torno de uma tendência determinística. Desta forma, toda a informação fornecida na parte inferior da tabela deve ser desprezada, isto é, deve usar-se apenas a informação produzida pelas equações do tipo da (11).

O quadro contém informação suficiente para usar a estratégia t -sig, com $k_{max} = 4$. Como essa estratégia será complementada com testes para detecção de autocorrelação, formalizem-se as respectivas hipóteses:

$$H_{0k} : v_t s \text{ sem autocorrelação de } 1^{\text{a}} \text{ ordem vs. } H_{1k} : \text{não } H_0,$$

sendo óbvio que não se deve usar a estatística d de Durbin-Watson, e

$$H'_{0k} : v_t s \text{ não autocorrelacionados até à ordem 4 vs. } H'_{1k} : \text{não } H_0.$$

Iniciando a análise com $k = k_{max} = 4$, constata-se que não se encontram indícios significativos da presença de autocorrelação: o valor- p da estatística $h - alt$ é de 0.61, e o da BG(4) é de 0.41, não se rejeitando as hipóteses nulas ao nível usual (de 5%). Desta forma, as inferências a efectuar para a respectiva equação parecem ser válidas.

Os testes a empregar no processo de simplificação sequencial são

$$H''_{0k} : \gamma_k = 0 \text{ vs. } H''_{1k} : \gamma_k \neq 0,$$

e para $k = 4$ tem-se $t_4 = 1.04$. Usando a aproximação normal para a distribuição das estatísticas de teste e $\alpha = 0.05$, tem-se

$$RC = \{t_k : |t_k| \geq 1.96\}.$$

Desta forma, a hipótese nula não é rejeitada e tenta-se simplificar a equação eliminando o desfasamento de ordem 4.

Para $k = 3$, o primeiro passo consiste em validar a exclusão de ΔU_{t-4} , testando a presença de autocorrelação nos erros da nova equação. Ora, como os valores- p das estatísticas $h-alt$ e $BG(4)$ são, respectivamente, 0.30 e 0.25, não se encontram sintomas de autocorrelação, pelo que a simplificação efectuada parece ser aceitável.

A tentativa de simplificação para $k = 2$ também não é rejeitada pois $t_3 = 0.85 \notin RC$. Desta forma, sobe-se mais uma linha no quadro e tenta-se avaliar se a simplificação não originou o aparecimento de autocorrelação nos erros da nova equação. Não parece ser esse o caso pois, mais uma vez, as estatísticas $h-alt$ e $BG(4)$ são insignificantes aos níveis habituais: a primeira com um valor- p de 0.40 e a segunda com 0.21. Todavia, a tentativa de simplificação seguinte, para $k = 1$, é claramente rejeitada: $t_2 = -2.96$, o que leva a rejeitar H''_{02} . Desta forma, o k seleccionado é $k(= \hat{k}) = 2$, produzindo $ADF(2) = \tau_c = -3.20$. Usando $\alpha = 0.05$, a região crítica vem dada por

$$RC = \{\tau_c : \tau_c < -2.86\}.$$

Por conseguinte, a hipótese nula é rejeitada. Embora os valores críticos tenham apenas validade assintótica (e a amostra é bastante pequena), note-se que mesmo que se usasse $\alpha = 0.025$ a nula de raiz unitária continuaria a ser rejeitada. Desta forma, a evidência contra a hipótese de raiz unitária parece ser bastante forte. Doutra forma, a série parece poder ser melhor aproximada por um modelo estacionário em torno de uma constante, não parecendo ser necessário diferenciá-la.

Finalmente, note-se ainda que se tivéssemos empregue a equação que minimiza a estatística AIC ter-se-ia obtido $k = 3$ e, conseqüentemente, $\tau_c = -2.44$ não permitiria rejeitar a hipótese de raiz unitária. Tal como neste, em muitos casos empíricos a escolha do k pode ser crucial para a decisão a tomar.